

## 中国語における述語動詞を中心とした 帰納的推論の計算モデルの構築

### The Construction of Computational Model for Inductive Reasoning in Chinese with a focus on predicate verb

張寓杰<sup>1</sup>, 張亜麗<sup>2</sup>, 寺井あすか<sup>3</sup>, 王婉瑩<sup>2</sup>, 菊地賢一<sup>1</sup>, 中川正宣<sup>4</sup>  
Yujie Zhang, Yali Zhang, Asuka Terai, Wanying Wang, Kenichi Kikuchi, Masanori Nakagawa

<sup>1</sup>東邦大学理学部情報科学科, <sup>2</sup>清華大学外国語学部日本語学科

<sup>3</sup>公立はこだて未来大学システム情報科学部 <sup>4</sup>大妻女子大学人間生活文化研究所

Toho University, Tsinghua University, Future University Hakodate, Otsuma Women's University

zhang@sci.toho-u.ac.jp

#### Abstract

In previous studies, we constructed a computational model of inductive reasoning based on the probabilistic concept structure estimated by the statistical analysis of large scale Chinese language data (e.g. Zhang, et al. 2013). This model mainly depended on similarities of nouns between premises and the conclusion. However, Heit and Rubinstein (1994) showed that not only nouns but also predicates affect the inductive reasoning.

In this study, adding to the previous model, the relationships between subject (nouns) and predicate (verbs), predicate (verbs) and objects (nouns) were included in the model based on the probabilistic concept structure estimated by the statistical analysis of large scale Chinese language data. Then, the validity of the model was verified using the psychological experiment.

Furthermore, for the comparison of simulation results between the previous model and the present one, we input the same predicate (verbs) and nouns in the both models. The results show that erroneous outputs of the previous model are improved successfully in the present one.

**Keywords — Inductive Reasoning, Statistical Language Analysis, Computational Model**

#### 1. 研究背景と研究目的

帰納的推論とは、初期の観察や命題に対して意味情報を増加させる結論を導く思考である。つまり、いくつかの個別知識から、一般法則を導き出す推論を意味する。

帰納的推論の心理学的メカニズムを説明するために、今までにさまざまな計算モデルが提案されてきた (Rips, 1975; Osherson, 1990; Sloman, 1993; Sanjana, 2002)。しかし、これらのモデルは、全て共通して、非常に限られた知識領域のみを対象とした帰納的推論以外は検証していないという問題点を含んでいる。心理実験の評定だけで、人間の持つ膨大な知識を構成する数えきれない概念間における類似度や特徴の関連強度を推定することは極めて困難である。坂本(Sakamoto

& Nakagawa 2007, 2008, 2010)は、以上の問題点に対し、日本語の大規模言語データの統計解析に基づいて確率言語知識構造を構成し、広い知識領域での帰納的推論の計算モデルを構築した。さらに、張他(2013)は中国語の帰納的推論の計算モデルを構築して、このモデルの日本語以外での有用性を明らかにし、日本語と中国語における帰納的推論の計算モデルのシミュレーション結果を比較している。

一方、Heit & Rubinstein (1994)は述語の帰納的推論への影響を明らかにするため、以下の心理実験を行った。

#### 解剖学的な特性

前提：鷹は二つの房のある肝臓を持っている。

結論：鶏は二つの房のある肝臓を持っている。 57.5%

虎は二つの房のある肝臓を持っている。 40.1%

#### 生活習慣に関する特性

前提：鷹は夜間の食事を好む。

結論：鶏は夜間の食事を好む。 42.2%

虎は夜間の食事を好む。 49.0%

たとえば、解剖学的な特性を述語として、「タカはふたつの房のある肝臓を持っている」という前提に対して、ニワトリは同じような特性をもっているという結論のもっともらしさを判断するとき、人間の判断の結果は 57.5%である。また、トラはこの特性を持っているという結論のもっともらしさはニワトリより低く、40%である。しかし、述語の部分を生生活習慣に関する特性に変えて、「タカは夜に食事を好む」とすると、トラは同じような特性をもっているという結論のもっともらしさはニワトリより高くなる。

この場合、「二つの房の肝臓をもっている」という特

性は生態学的な特性であり、タカとニワトリは同じ生態学的カテゴリーである鳥類に属するので、その類似性はタカとトラの類似性より大きくなる。しかし、述語を生活習慣に変えると、タカとニワトリの生活習慣に関する類似性はタカとトラより小さいため、トラが同じ特性をもっているという結論のもっともらしさはニワトリより大きくなると考えることができる。

以上のように、Heit & Rubinstein (1994)は、同じ前提条件の名詞に対して、述語動詞を変えると、人間の帰納的推論の判断が変わるという実験結果を示している。しかし、前記の坂本や張の計算モデルでは帰納的推論における前提や結論に含まれる名詞部分しか考慮しておらず、前提や結論の述語が帰納的推論に与える影響は考慮していない。

本研究の目的は、このような先行研究の問題点を解消するため、述語動詞の影響を考慮した帰納的推論の中国語における計算モデルを構成し、心理学実験を行い、計算モデルの妥当性を検証することである。さらに本研究では、本モデルと張他(2013)で構築された中国語の計算モデルとを比較し、本モデルが先行モデルより妥当性が高いことを実証する。

## 2. 研究方法

本研究では、張他(2013)の計算モデルを修正し、結論のもっともらしさを計算する際、結論における名詞と動詞間の類似性を加えて、以下のような新たなモデルを構成する。

ここで  $v(N_i^c)$  は結論  $(N_i^c)$  の尤もらしさの値、 $P(c_k | N_i^c)$ ,  $P(c_k | N_j^+)$ ,  $P(c_k | V)$  は各々結論  $(N_i^c)$ , 前提事例の名詞  $(N_j^+)$ , 述語動詞  $(V)$  が与えられたときの潜在クラス  $(c_k)$  の条件付確率を示している。これらの条件付確率は中国語の大規模言語データの統計解析に基づきあらかじめ推定されている(張他 2013 参照)。また、 $d_{ij}^+$  は、各名詞の潜在意味クラスの条件付確率  $P(c_k | N_i^c)$ ,  $P(c_k | N_j^+)$  に基づいて計算される結論  $N_i^c$  と前提事例  $N_j^+$  の距離であり、 $d_i^v$  は、 $P(c_k | N_i^c)$ ,  $P(c_k | V)$  に基づいて計算される結論  $N_i^c$  と述語動詞  $V$  の距離である。 $m$  は距離の計算に用いる潜在意味クラスの数で、対象が潜在意味クラスへのメンバーシップ値からなる  $m$  次元の特徴空間において距離が計算される。 $SIM_+(N_i^c)$  と  $SIM_v(N_i^c)$  は各々、結論と前提事例の名詞、結論と述語動詞の類似性の大きさを表す。 $a$  と  $b$  は、前提名詞と述語動詞の重みづけパラメータで、 $B$  は  $d_{ij}^+$  と  $d_i^v$  の変化が類似性の変化に影響する程度を表す相対的感度を示している。

$$v(N_i^c) = aSIM_+(N_i^c) + bSIM_v(N_i^c) + h \quad (1)$$

$$SIM_+(N_i^c) = \sum_j^{n^+} e^{-\beta d_{ij}^+} \quad (2)$$

$$SIM_v(N_i^c) = e^{-\beta d_i^v} \quad (3)$$

$$d_{ij}^+ = \sqrt{\sum_k^m (P(c_k | N_i^c) - P(c_k | N_j^+))^2} \quad (4)$$

$$d_i^v = \sqrt{\sum_k^m (P(c_k | N_i^c) - P(c_k | V))^2} \quad (5)$$

表1 シミュレーション課題

主語プラス動詞 (S+V)		
	名詞	動詞
課題1	指针 (計測器の針) 树枝 (木の枝)	巻曲 (曲がる)
課題2	指针 (計測器の針) 树枝 (木の枝)	摆动 (揺れる)
課題3	水管 (水道管) 河流 (川)	断裂 (断裂する)
課題4	业绩 (業績) 成绩 (成績)	达标 (標準に達する)
課題5	业绩 (業績) 成绩 (成績)	超群 (ずば抜ける)
課題6	书籍 (書籍) 商品 (商品)	面世 (出てくる)
動詞プラス目的語 (V+O)		
課題7	特点 (特点) 风格 (風格)	具有 (持つ)
課題8	时间 (時間) 机会 (機会)	争取 (勝ち取る)
課題9	运动员 (スポーツ選手) 队员 (メンバー)	带领 (率いる)

この方法で構築した計算モデルに基づいてシミュレーションを行った。シミュレーションで入力した名詞と動詞の項目は以上の表1に示されるように述語動詞として使われる動詞と名詞の組み合わせは、「主語+動詞 (S+V)」と「動詞+目的語 (V+O)」の二種類があるので、今回のシミュレーションでは両方を用いた。

### 3. 結果と考察

シミュレーションで用いた課題は、「主語+動詞 (S+V)」の課題6個、「動詞+目的語 (V+O)」の課題3個、合計9個の課題である(表1参照)。

表2 主語+動詞 (S+V) のシミュレーション課題の例

名詞	日本語訳	動詞	日本語訳
指针	計測器の針	卷曲	曲がる
树枝	木の枝		

表3 先行研究のモデルの出力結果

結論	日本語訳	尤もらしさ
树枝	木の枝	1.519
指针	計測器の針	1.519
低産田	収穫量の低い田畑	1.334
鉄鏈	鉄製のチェーン	1.325
附加費	ほかの費用	1.318
針頭	注射針	1.310
核裁軍	核軍縮	1.309
党群	党と大衆	1.307
树干	木の幹	1.293
二拇指	人差し指	1.286

表4 本研究のモデルの出力結果

結論	日本語訳	尤もらしさ
指针	計測器の針	2.942
树枝	木の枝	2.908
树干	木の幹	2.866
樹根	木の根	2.864
軀干	胴	2.848
二拇指	人差し指	2.835
腰身	ウエスト	2.828
低産田	収穫量の低い田畑	2.791
附加費	ほかの費用	2.776
帘布	カーテン	2.775

以上では、本研究のモデルと先行研究の張他(2013)のモデルのシミュレーション結果を比較する。以上の表3, 4の例は、先行モデルと本モデルの各々に、表2の主語+動詞 (S+V) 課題の2つの名詞と一つの動詞を入力した場合の出力結果である(ただし尤もらしさの値の上位10個の単語だけを抽出)。

表3の先行研究のモデルの出力結果では、動詞「卷曲(曲がる)」と組み合わせても意味の通じない名詞「低産田(収穫量の低い田畑)」、「附加費(ほかの費用)」、「核裁軍(核軍縮)」、「党群(党と大衆)」が上位に出力されていたが、表4の本研究のモデルの出力結果では、動詞「卷曲(曲がる)」と組み合わせても意味の通じる名詞の方が上位に出力されている。つまり、本研究のモデルのほうが全体的に妥当性が高いと考えられる。

一方、表6, 7は、先行モデルと本モデルの各々に、表5の動詞+目的語 (V+O) 課題の2つの名詞と一つの動詞を入力した場合の出力結果である(ただし尤もらしさの値の上位10個の単語だけを抽出)。

表6の先行研究のモデルの出力結果では、動詞「争取(勝ち取る)」と組み合わせても意味の通じない名詞「整流器(整流器)」、「纸张(紙)」、「家用电器(家電製品)」が上位に出力されていたが、表7の本研究のモデルの出力結果では、「公民権」、「果実」など、動詞「争取(勝ち取る)」と組み合わせても意味の通じる単語が出力されている。また、やはり動詞「争取(勝ち取る)」と組み合わせても意味の通じない名詞「纸张(紙)」の順位が本研究のモデルでは先行モデルより下がっている。つまり、本研究のモデルのほうが全体的に妥当性が高いと考えられる。

以下の図1はモデルのシミュレーション結果と心理学実験における評定平均値との相関係数が示されている。図1からわかるように、9課題共に本研究で提案したモデルの結果は先行研究のモデルの結果より相関係数が高く有意である。

これらの結果から本研究で構成した計算モデルが中国語の帰納的推論における述語動詞の影響をより正確に表現できているといえる。

今後、同じ手法で日本語の述語動詞を中心とした帰納的推論の計算モデルを構築し、モデルの妥当性を検証した上、本研究で構築した中国語の計算モデルとの比較を行う予定である。

表5 動詞+目的語 (V+O) のシミュレーション  
課題の例

名詞	日本語訳	動詞	日本語訳
時間	時間	争取	勝ち取る
機会	機会		

表6 先行研究のモデルの出力結果

結論	日本語訳	尤もらしさ
時間	時間	1.393
機会	機会	1.393
良机	好機	1.140
整流器	整流器	1.122
纸张	紙	1.121
贴息贷款	利子補給の貸付	1.116
贴息	利子	1.116
优秀奖	優秀な賞	1.114
家用电器	家電製品	1.071
外援	援助者	1.056

表7 本研究のモデルの出力結果

結論	日本語訳	尤もらしさ
特权	特権	2.675
果实	果実	2.659
公民权	公民権	2.629
機会	機会	2.528
决赛权	決勝戦の権力	2.513
优秀奖	優秀な賞	2.509
豁免权	治外法権	2.462
时间	時間	2.461
良机	好機	2.453
纸张	紙	2.315

参考文献

[1] Heit, E., & Rubinstein, J. (1994) "Similarity and property effects in inductive reasoning", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 411-422.

[2] Kameya, Y., & Sato, T. (2005) "Computation of probabilistic relationship between concepts and their attributes using a statistical analysis of Japanese corpora", *Proceedings of Symposium on Large-scale Knowledge Resources*, 65-68.

[3] Kayo Sakamoto, Asuka Terai, Masanori Nakagawa (2007) "Computational models of inductive reasoning using a statistical analysis of a Japanese corpus", *Cognitive Systems Research*, 8, 282-299.

[4] Kayo Sakamoto, Masanori Nakagawa (2008) "A Computational Model of Risk-Context-Dependent Inductive Reasoning Based on a Support Vector Machine", T. Tokunaga and A. Ortega (Eds.): LKR2008, LNAI 4938, Springer-Verlag Berlin Heidelberg, pp.295-309

[5] Kayo Sakamoto, Fang Xie, Masanori Nakagawa (2010) "Syntactic Dependency Analysis Reveals Semantic Concept Structure Underlying Inductive Reasoning: Towards a Domain-Inclusive Structure that Enables Context-Dependent Knowledge Selection", *Cognitive Studies*, Vol.17, No.1, 143-168.

[6] Osherson, D. N., Smith, E. E., Wilkie, O., López, A., and Shafir, E. (1990) "Category based induction", *Psychological Review*, 97, 185-200.

[7] Rips, L. J. (1975) "Inductive judgment about natural categories", *Journal of Verbal Learning and Verbal Behavior*, 14, 665-681.

[8] Sloman, S., A. (1993) "Feature based Induction", *Cognition*, 49, 67-96.

[9] Sanjana, N., & Tenenbaum, J. B. (2002) "Bayesian models of inductive generalization", In: Becker, S., Thrun, S., Obermayer, K (Eds.) *Advances in Neural Information Processing Systems*, 15. Cambridge, MA: MIT Press.

[10] 張寓杰, 寺井あすか, 董媛, 王月, 中川正宣, (2013) "日本語と中国語における帰納的推論の比較研究—言語統計解析に基づく計算モデルを用いて—", 『認知科学』, No. 20 Vol. 4, 439-469.

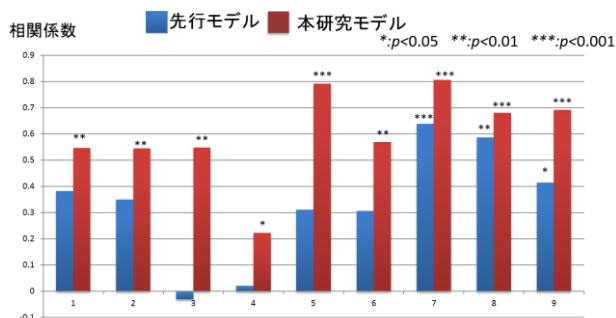


図1 先行モデルと本研究モデルに対してシミュレーション結果と心理学実験における評定平均値との相関係数の比較